WILL ENERGY STORAGE COME OFF THE BENCH IN ARGENTINA? ANALYZING RESIDUAL LOAD FROM RENEWABLE ENERGY SURPLUS GENERATION FOR 2026 SCENARIOS

Javier E. Navajas Ahumada, Argentina Secretary of Energy, +5491123975738, jnavajas@energia.gob.ar

Abstract

In an international context of low carbon energy transition, many countries have started deploying renewable power generation which has placed interest in the development of energy storage to harvest residual load. Argentina has recently set a 20% renewable electric energy consumption target by December 31th 2025. This study aims to estimate whether Argentina will produce residual load by 2026 assuming full deployment of renewable energy for three different demand scenarios. An energy demand forecasting model for Argentina was developed using a hybrid model (similar day method and SARIMA time series) based on historical hourly energy demand data of Argentina from 2007-2017, and macroeconimic data. To estimate inflexible generation, bottom-up aggregate generation models were developed for wind and solar energy. 2017-2018 inflexible generation data was used to build Monte Carlo simulations, regressions and time series models. 2008-2018 hydroelectric power generation data was also included in the residual load estimation. This study shows that by 2026 residual load is unlikely to generate surplus energy to be stored, at least, at a moderate scale.

Introduction

In an international context of low carbon energy transition set by the 2015 Paris Agreement, many countries have started deploying renewable power generation. An early example is the German '*Energiewende'* which states that renewable energy sources have to increase from roughly 10% by 2010 to 60% by 2050 [1]. Since renewable energies are intermittent in nature (eg. wind, solar, etc.), this has placed a significant interest in the development of energy storage to harvest residual load –surplus generation that exceeds demand- [2].
In the opposite direction Argentina has early developed energy storage facilities, specifically pumped-hydro power plants: Los Reyunos (1983) and Río Grande (1986) that provide a total of 974MW of installed capacity. On the other hand, according to Law Nº26.190 (later modified by Nº27.191) contribution of renewable energy in Argentina should reach 20% national electric energy consumption by December 31th 2025. In order to reach this goal the Argentine Government has released RenovAr 1, 1.5 and 2 tenders to install a total capacity of almost 4500 MW and is expected to develop further over the next years.
This work aims to predict whether renewable energy will produce residual load by 2026 and if there will rise a business opportunity for Argentina's sunk energy storage infrastructure to harvest renewable energy generation surplus.

Residual Load

As mentioned in the introduction, residual load lets evaluate surplus generation left for energy storage. This section discusses the definition of residual load and defines a methodology for estimation.
According to [3], residual load can be defined as the load *"... left after substracting those generators who have to produce electricity ('must run') and those that generate with (almost) no marginal costs (variable renewables like wind, solar and hydro)*". Although this definition sheds some light on what generation should be included in the calculation (wind and solar by definition fall into this category), it is not extensive to all types of power generation sources and unclear whether all hydroelectric capacity should be included in this calculation.
In Argentina, renewable projects have priority to dispatch energy (this includes wind, solar, biomass, biogas, mini-hydro projects) and are intrinsically inflexible. Also, nuclear energy can be considered 'must run' given that, by default, it operates at maximum capacity given its almost zero marginal cost.
Run-of-river hydroelectric generation are most of the time inflexible given that these projects usually have little or no energy storage. Impoundment hydropower plants, however, should only be considered if these are restricted to store energy (eg. restricted by operations). This topic will be later discussed.

Taking the aspects discussed above into consideration, equation (I) provides a clear definition on how to estimate residual load at a given hour (h):

$$RL^h = TL^h - G^h_{wind} - G^h_{solar} - G^h_{nuclear} - G^h_{BM+BG} - G^h_{hydro(RoR)} - G^h_{hydro(I-Inf)} \quad \text{(I)}$$

where RL stands for residual load, TL is the total system load, BM+BG stands for biomass and bio gas, RoR=Run-of-the-River, I-Inf=Inflexible Impoundment. If RL=0, inflexible generation meets demand. On the other hand, if RL<0 generation exceeds demand which could be potentially used for energy storage.
The next sections will provide, then, an estimation on demand and generation by 2026 discussing how to calculate all these terms.

Energy Demand Modelling

Database

The data used to forecast the aggregated demand was a 2007-2018 hourly demand data of the SADI (Sistema Argentino de Interconexión) which is Argentina's main electricity transmission network. Although information of previous years was also available, SADI integration to the Patagonia tranmission Subnetwork SIP (Sistema Interconectado Patagónico) occurred in March 2006.
This study also assumes for the 2026 forecast that the integration of other interconnections -eg. Sistema Fueguino de Subtransmisión which could mean an increase in demand of about ~100MW- is unlikely to occur in the following years given that the SADI Sur has already committed its transmission capacity. Penetration of electric vehicle technologies were also disregarded for this study.

Electricity Load Model Building

The model choose to forecast electricity demand is based on the approach described by Berk [4]. Let C_t be the electricity load of the SADI interconnection at time t, we describe the load by the following equation:

$$logC_t = D_t + u(G_t^*) + R_t \quad \text{(II)}$$

where D_t denotes a deterministic load forecast, $u(G_t^*)$ a function of the residual grid load and Rt the residual time series. Logarithm was applied to the electricity load to reduce skewness. It is important to outline that the deterministic load model must include trend –i.e predictors of long-term economic development- and seasonal effects –eg. season, monthly, weekly, etc.- as proposed in [5,6]. The following sections will give a detailed explanations of all model components.

Data Cleansing & Outlier Cleaning

First, data cleansing was performed to detect corrupt data (eg. zero values) using a hampel filter with 120-hr and 5 times the standard deviation (replacing the mean for the median for robustness). Points were checked before removal and replaced using average of neighboring values.

Before training the deterministic load forecast model, outliers were removed from the electricity load database using a hampel filter. The same procedure as described above was chosen with 120-hr and 3 times the standard deviation. Nevertheless, it is important to outline that the original (cleaned) dataset was used later to compute and analyze residual loads.

Deterministic Demand Forecast Modelling

The deterministic load forecast was modelled applying a similar-day method to the historical data set. Other variables such as population, GDP and Wholesale electricity prices were included to model economic cyclic and long term trends. Based on this approach, a set of explanatory variables (Table I) was developed to explain trend and seasonal changes in the aggregated electricity load (response variable). All variables were normalized to avoid distortion differences.

ϑ_1: Monday

ϑ_2: Tuesday

ϑ_3: Thursday

ϑ_4: Friday

ϑ_5: Saturday

ϑ_6: Sunday

ϑ_7: January 1st

ϑ_8: January 2nd

ϑ_9: December 25th

ϑ_{10}: December 31st Eve (12-24hs)

ϑ_{11}: December 24th Eve (12-24hs)

ϑ_{12-22}: Month i (i=12=January, i=13=February, i=14=April...,i=22=December)

ϑ_{23}: Public Holiday/Bridge Day

ϑ_{24}: Winter Holiday Period

ϑ_{25-27}: Season i (i=25=Summer, i=26=Winter, i=27=Spring)

ϑ_{28}: Population (daily linearized from INDEC census)

ϑ_{29}: GDP (USD, quarterly, deseasonalized)

ϑ_{30}: Wholesale Electricity Price for Large Users (USD,monthly)

Table I. Overview of variables for deterministic load forecasting.
ϑ_{1-27} corresponds to binary variables. ϑ_{28-30} are continuous variables.

It is important to outline that ϑ_{29} corresponds to the wholesale electricity price for large users instead of the average wholesale electricity price. This was considered as being more representative given the discriminatory nature of the wholesale electricity market that still prevails in Argentina, where the gap between industrial and residential price is significant.

A separate deterministic forecast for each hour was preferred over taking the time series as a whole set given initial assessment comparing linear regressions gave significant lower errors.

Next, the performance of different machine learning algorithms (multiple linear regression, robust regression, regression trees and LSBoost/Bag ensemble learning) was studied in order to select the most adequate model. For this purpose the dataset was divided in training test (2007-2016 historical data) and test set (2017 historical date).

Parameter tuning for the regression tree was conducted using cross validation over the training set. Regression Ensemble optimization was optimized using Matlab automatic hyperparamter optimization that minimizes five-fold cross-validation. Robust Regression optimization was conducted choosing amongst different weight functions (‘Andrews’, ‘Bisquare’ and ‘Cauchy’) and tuning the weight constant. Mean squared error (MSE) was chosen as the error estimator for conducting these analysis. The results are summarized in Figure I.

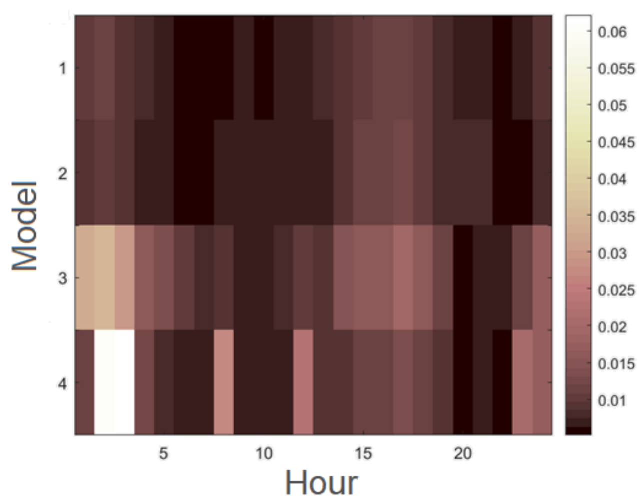

Figure I. MSE value comparison (colorscale) for each algorithm (1=Linear Regression, 2=Robust Regression, 3=Regression Trees, 4= Regression Ensemble) vs hourly data (N=24).

From the results shown, it can be concluded that Linear and Robust Regression had similar performance throughout each hour. Regression Trees and Regression Ensemble, on the other hand, had variable performance. Based on this analysis, it was decided to use Robust Regression to build our model.

Table II summarizes the results from fitting robust regression model for each hour. In general terms, curve adjustment gives good results (R^2 ranges from 0.83 to 0.92) and RMSE values are moderate as well (from $4.44E^{-02}$ to $6.11E^{-02}$). Figure III shows an example of how the deterministic model fits the hourly data used for the regression.

Hour	R^2 Adjusted	RMSE	Hour	R^2 Adjusted	RMSE
0	0.84	5.39E-02	12	0.89	5.22E-02
1	0.95	3.13E-02	13	0.88	5.61E-02
2	0.85	5.26E-02	14	0.88	5.81E-02
3	0.84	5.18E-02	15	0.88	5.92E-02
4	0.85	4.95E-02	16	0.87	6.10E-02
5	0.84	4.81E-02	17	0.86	6.11E-02
6	0.85	5.00E-02	18	0.84	5.92E-02
7	0.9	4.89E-02	19	0.83	5.40E-02
8	0.92	4.89E-02	20	0.83	4.69E-02
9	0.92	4.93E-02	21	0.84	4.44E-02
10	0.91	4.89E-02	22	0.83	4.83E-02
11	0.9	5.01E-02	23	0.83	5.32E-02

Table II. Summary of Robust fit results for each hour regression (N=24)

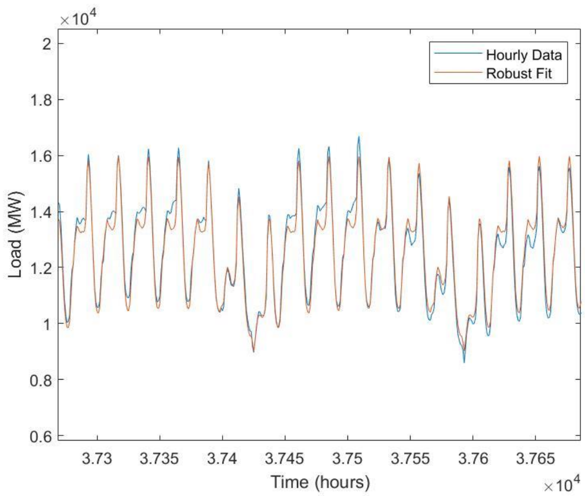

Figure II. Robust Fit vs. Hourly Data (Regression Data, random sample)

Residual Demand Forecast Modelling

The residual demand results from substracting the determinsitic demand to the original (cleaned) dataset. Figure III shows the residual load that will be modelled using time series.

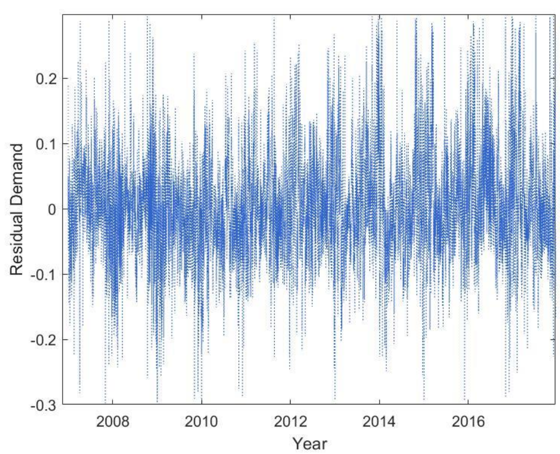

Figure III. Residual Load Time Series.

A Dickey and Fuller test was performed to check for stationarity. The test indicated that the hypothesis is TRUE so we can reject the unit root hypothesis with a significance level of pvalue=10^{-3}. Therefore, we can assume that the residual grid demand is stationary.

Next, the sample partial autocorrelation function was obtained to study potential time series models (see Figure IV). This graphs show a strong autocorrelation with the 1st lag and a clear 24 hour seasonality. Therefore, a SARIMA (p, 0, q)×(P, 0,Q)24 model were built in accordance to the mentioned autocorrelation structure.

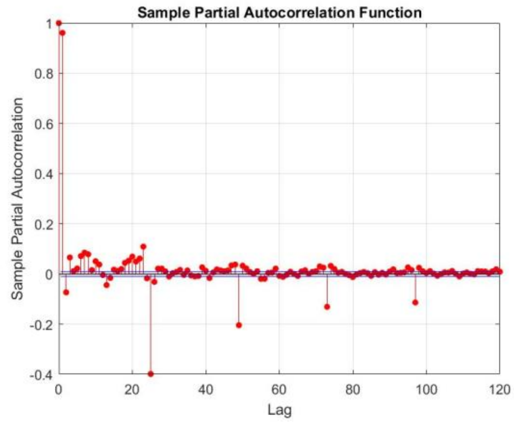

Figure IV. Sample Partial Autocorrelation Function

In order to determine and contrast the performance of the model parameters, the Akaike Information Criterion (AICC) was calculated for each case. The values were calculated for $p, P, q, Q \in \{0,1,2,3\}$ which gives a total of 4^4=256 models. Table III provides a ranking of the models that shown the best performance.

Ranking (#)	p	q	P	Q	AICC
1	3	3	3	3	-549804
2	3	3	3	2	-549801
3	3	2	3	3	-549790
4	3	3	1	3	-549789
5	3	2	2	3	-549787

Table III. Performance ranking for time series models (p, 0, q)×(P, 0,Q)24 using AICC

Given that the difference between these values is not considerably significant the model with less parameters $(3,0,2)\times(2,0,3)_{24}$ was chosen to describe the residual grid load. The model parameters obtained were the following:

```
  Description: "ARMA Error Model Seasonally Integrated (t Distribution)"
 Distribution: Name = "t", DoF = 3.10448
    Intercept: 0
         Beta: [1×0]
            P: 72
            Q: 72
           AR: {1.20736 -0.206001 -0.0281026 0.00307452 -0.00176962} at lags [1 2 3 24 48]
          SAR: {}
           MA: {0.00463636 -0.00680636 -0.598456 -0.127095 -0.0532411} at lags [1 2 24 48 72]
          SMA: {}
  Seasonality: 24
     Variance: 0.000175045
```

Figure V. Model parameters for the SARIMA model.

In order to check if the model captured the autocorrelation effects, the partial autocorrelation function of the time series residuals was obtained as shown in Figure VI. This graphs shows that most autocorrelation was removed to a significant level (confidence interval $\pm 1.96\sqrt{M}$ with M=96360 hours).

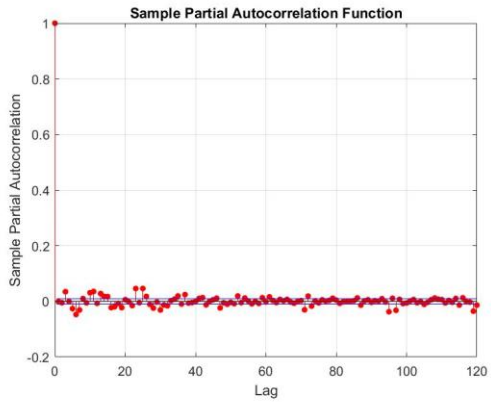

Figure VI. Sample Partial Autocorrelation for the time series residuals

Last, an analysis of the time series residuals was also conducted to verify the distribution of the model innovations. As shown in Figure VII, it can be seen that these follow a t-Student distribution with ν=3.1, which was considered for the forecasting.

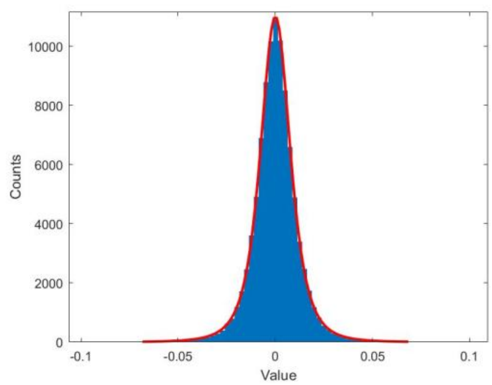

Figure VII. Time series residuals histogram. The results show a t-student distribution with ν=3.1

2026 Scenario Demand Forecast Modelling

In this section, a brief description of three different scenarios will be provided. This data will serve to build the variables for the deterministic regression described in Table I. In particular, variables $\vartheta_{1\text{-}27}$ were built taking the holiday scheme of 2017 whereas $\vartheta_{28\text{-}30}$ requires a more sophisticated analysis.
Population (ϑ_{28}) was estimated projecting last population census (INDEC 2010) assuming a compound annual grow rate of 1.14% which was the growth for the census year. Variables $\vartheta_{29\text{-}30}$ (GDP and Wholesale Electricity price) were estimated under three different scenarios of Argentina development (Low, Base and High). These are summarized in Table IV.

Scenario	GDP Growth	Wholesale Electricity Price for Large Users (USD) in 2026
Low	1.5%/yr.	100
Base	2.5%/yr.	90
High	3.5%/yr.	70

Table IV. Argentina Development Scenarios for 2026 Scenarios

Wind Generation Modelling

Wind Generation Data Analysis

The first part of this section aims to study the effects of spatial correlation of wind power generation across Argentina. This was conducted using a similar approach as described by Malvaldi [7]. The data gathered for this analysis includes the generation data of wind power plants in Argentina for the years 2017-2018.
The power generation information was carefully filtered (eg. testing periods of wind power were removed, scheduled maintenance) and normalized (setting capacity power to 1). Geographic coordinates of the wind power generation plants were used to calculate distance between power plants.
Figure VIII shows correlation coefficients for wind power plants pairs versus the distance using an exponential function as described again by Malvaldi:

$$\tau = a\, e^{b*d} \tag{III}$$

where τ is the correlation, d is the distance, and a,b are coefficients. In this study, coefficient **a** was set to 1 which can be interpreted as τ=1 when d=0. Robust fitting was used using Andrews weight. Coefficient points with distance over 1000km with negative correlation was assumed to be 0 (not statistically significant). The value of the coefficient using this approach gives $b=-4.8*10^{-3}$. This Figure shows a strong correlation for short distances (up to ~0.5 for distance around ~150km).

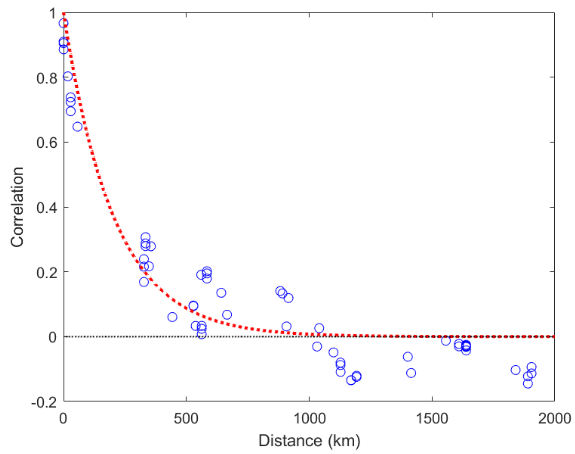

Figure VIII. Correlation vs. Distance (in kilometers) plot.
Data was fitted as described in equation (III) $\tau = e^{4.8*10^{-3}*d}$

Next, correlograms for each pair of wind power plants were constructed. Figure IX shows two correlograms. Figure IXa shows one with a wind park pair at a close distance (adjacent parks) and hence high correlation (τ=0.97) versus Figure IXb at a longer distance (d=57km) and lower correlation (τ=0.65).

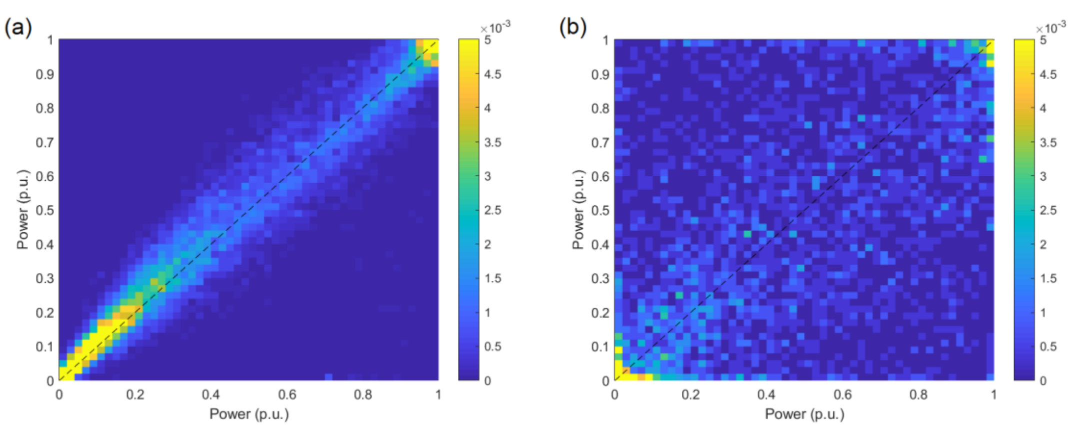

Figure IX. a) Rawson I vs. Rawson II Wind Power Plant Correlogram (τ=0.97) b) Corti vs. La Castellana Wind Power Plant Correlogram (τ=0.65). Normalized probability shown in colorscale. Max color scale was set to 0.005 to show dispersion.

Aggregated Wind Power Energy Simulation

The second part of this section briefly describes the procedure used to run aggregate wind power simulations. A bottom-up approach was selected using the wind power generation data described above and infrastructure metadata (coordinates and installed capacity) of the wind park projects that should be operational over the next years (Renovar 1, 1.5, 2 and MATER).

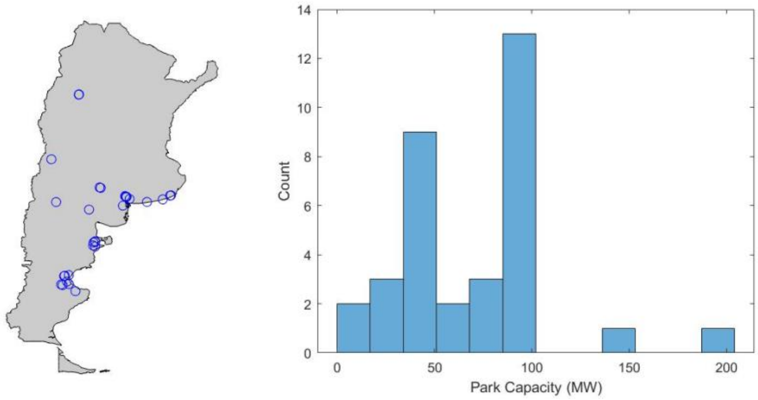

Figure X. Distribution of Res.202, Renovar 1, 1.5 and 2 wind park projects

Wind power park infrastructure by 2025 was extrapolated assuming installed plus to be operational wind power plants will be available by that date. Given that solar capacity by 2026 will exceed the former projects, additional power plants were added randomly in the main wind areas following the same distribution as the Renovar and MATER projects.

A Monte Carlo approach was selected to build aggregate wind power generation. The main concept behind this approach is to: 1) use 2017-2018 wind power generation information of operational projects as seeds 2) generate sample data running Monte Carlo simulations using artificial correlograms built from empirical correlograms described in the previous section. A detailed explanation of how artificial correlograms were built are described in the next section.

The pseudo code used for the simulation is the following:

i) Choose a park (that is not a seed)
ii) Find closest seed
iii) Calculate distance using geographic coordinates
iv) Estimate correlation using equation (III)
v) Build artificial correlogram
vi) Run Monte Carlo trials for each hour of the year (8760 trials) using generation data from (ii) and correlogram (v).
vii) Repeat. If correlation is lower than 0.5 (no close seeds are available), then include simulated data from (vi) to run as seeds.

Figure XI shows the result of a simulation using the procedure described in this section.

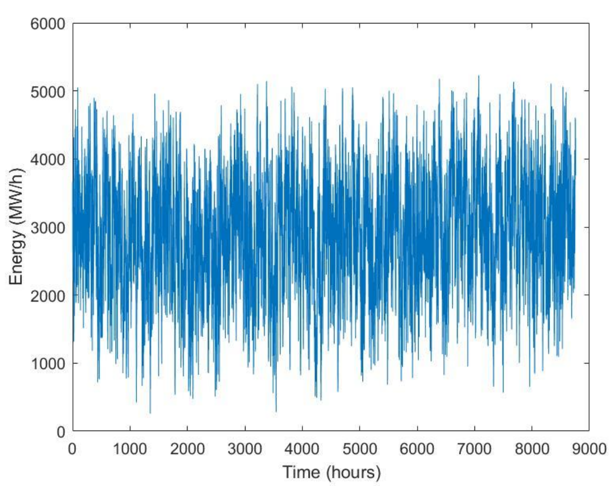

Figure XI. 6,000MW wind power installed capacity energy simulation for 8760h (1year)

Building Correlograms for Monte Carlo Simulation

In this section, a brief explanation of how to build artificial correlograms will be given. The main purpose behind this procedure is to be able to build a representative correlogram using only the correlation as the input.
For a given correlation τ , artificial correlogram C is built using a Gaussian kernel function that uses the empirical correlograms described above (C_i') associated to their respective correlation coefficients (τ_i'):

$$C(\tau) = \sum_{i}^{n} w_i(\tau)\ C_i' \tag{IV}$$

where wi are the weights associated to each of the n empirical correlograms C_i' which are calculated as shown below:

$$w_i(\tau) = \frac{e^{-\beta\|\tau_i'-\tau\|^2}}{\sum_i e^{-\beta\|\tau_i'-\tau\|^2}} \tag{V}$$

In this equation, β is a decay constant which was set arbitrarily to β=100. It can be seen, for a given β, the closer the correlation τ to the empirical correlation τ_i', the higher the weight. Figure XII shows an example of an artificial correlogram built with this procedure for τ=0.8

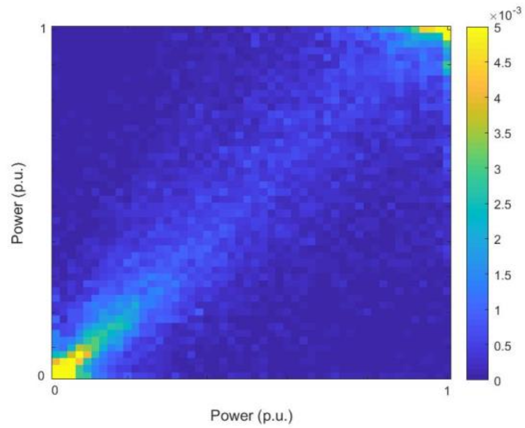

Figure XII. Artificial correlogram built for τ=0.8. Normalized probability shown in colorscale. Max color scale was set to 0.005 to show dispersion.

Solar Energy Modelling

Deterministic Solar Energy Modelling

The approach to generate simulated solar photovoltaic power plant data is based on two components: a deterministic component which depends on the performance of the PV systems vs. a stochastic component produced -mainly- by cloud shadowing.
The list of variables that should be considered to model the performance of PV plants is vast: solar radiation (direct and diffuse) at a specific power plant, temperature, the technology of the PV modules (eg. Si monocrystalline has a better performance than Si polycristalline), array orientation (eg. two-axis tracking gives higher yields than single-axis though the costs associated to the former are higher), array shading (eg. row to row, objects or horizon shadings can reduce energy output) and electrical losses (eg. wire losses, inverter and transformation losses) to mention just a few dimensions. On the other hand, strong cloud shadowing models usually account for temporal and spatial correlation based on weather data.
Given that we are interested in long term aggregated solar energy from PV plants and the information described above is limited –specially for 2026-, a simplified approach was proposed for modelling. This consists of using a similar approach that was used for demand modelling by relating energy output to direct radiation as described next:

$$G_t = D_t + u(G_t^*) + R_t \tag{VI}$$

where, in this case, D_t denotes a deterministic solar energy forecast, $u(G_t^*)$ a function of the residual solar energy and R_t the residual time series.

The procedure to gather information for the model was the following:

1) Build direct solar radiation data based on geographic coordinates.
2) Plot Energy Output vs. Direct Radiation for a given PV plant
3) Remove outliers (cloud shadowing) for deterministic modelling
4) Cluster performance of solar power plants
5) Fit Deterministic Regression for each cluster
6) Calculate residuals from deterministic model fitting
7) Use time series for solar energy residuals (cloud shadowing)

The data gathered for this analysis corresponds to the hourly generation data of solar power plants in Argentina for the years 2017-2018. Given the scarcity of information of solar parks in operation (191MW according to Jan-19 CAMMESA month report), step (4) was simplified to two categories: small plant (less than 20MW) and large plants (more than 20MW).
Direct Solar Radiation data was obtained from Suncycle software -available at UCSD SCRIPPS Institute of Oceanography webpage [8]- that gathers information from the US. Naval Observatory. Figure XIII shows an example of the direct solar radiation calculated for a given year.

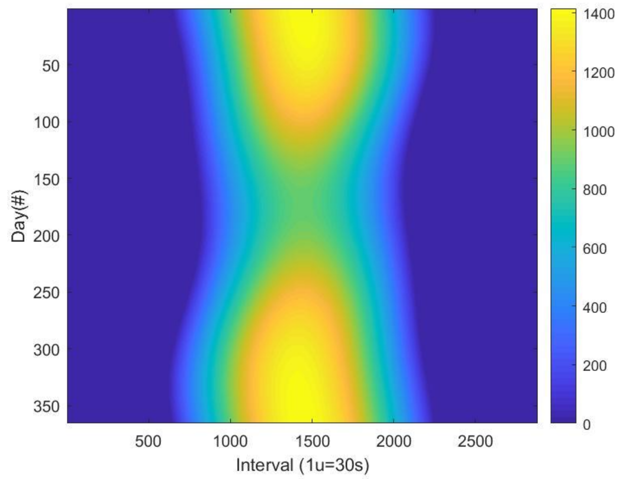

Figure XIII. Direct Solar Radiation (no-sky) values –shown in colorscale- calculated for coordinates (-24.14,-66.91) during year 2026.

Figure XIV shows a plot of sqrt(Power) vs. Solar Radiation for a given solar power plant. Lines connect hour-to-hour datapoints. Outliers below the curve were filtered for each power plant.

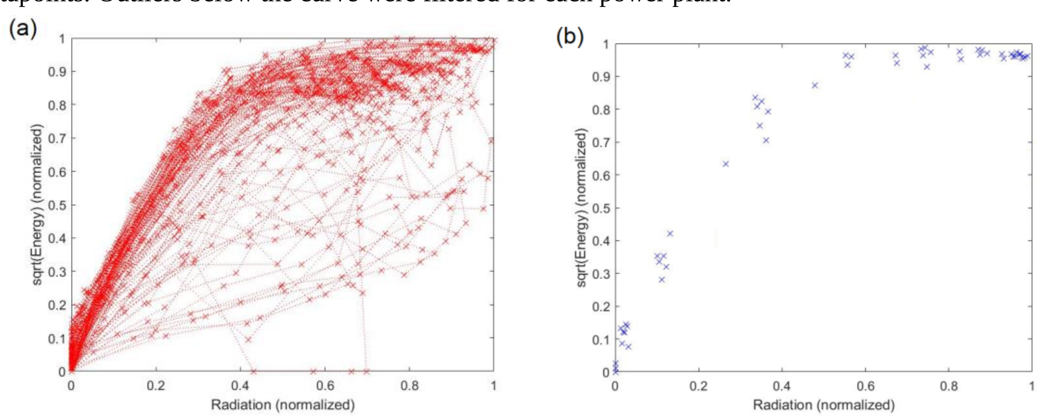

Figure XIV. (a) sqrt(Power) vs. Direct Solar Radiaton for a PV (raw data). Energy was normalized by the plant capacity. Solar Radiation was normalized by its maximum value (b) example of (a) for a given week with outliers removed for fitting.

Next, regressions were built for both categories of power plants using as predictors the variables summarized in Table V. It is worth to outline that although direct radiation captures seasonality effects of power output, as described at the beginning of this section, other variables such as temperature do also have seasonal effects that affect performance which are tried to be captured by adding “week” binary variables.

$\boldsymbol{\vartheta_{1\text{-}53}}$: Week i (i=1,...,53) (binary variables)
$\boldsymbol{\vartheta_{54}}$: Direct Radiation
$\boldsymbol{\vartheta_{55}}$: Direct Radiation^2
$\boldsymbol{\vartheta_{56}}$: Direct Radiation^3

Table V. Overview of variables for deterministic solar energy forecasting.
$\boldsymbol{\vartheta_{1\text{-}53}}$ corresponds to binary variables. $\boldsymbol{\vartheta_{54\text{-}56}}$ are continuous variables.

Figure XV shows an example of how a deterministic model (for >20MW plant) fits the hourly data used for the regression.

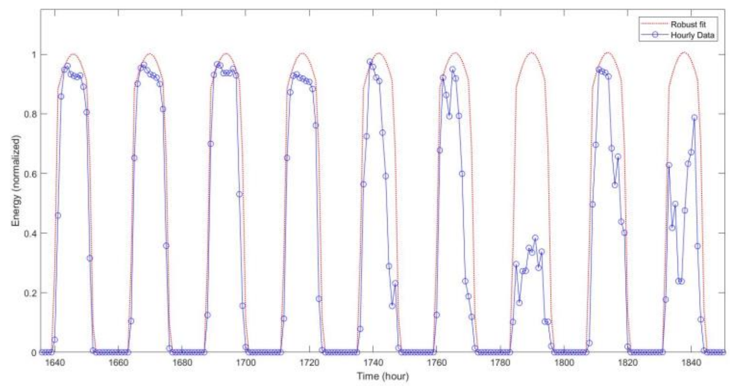

Figure XV. Robust Fit vs. Hourly Data (Regression Data, random sample)

Stochastic Solar Energy Modelling

The residual solar energy results from substracting the deterministic model to the original hourly data. Figure XVI shows the residual that results from this operation. Time periods where there was not energy output (i.e. after sundown before sunrise) were estimated using a sunrise and sundown estimation function available with Suncycle. A binary variable (δ=1= After sunrise, before sundown; δ=0= After sundown, before sunrise) was added to model the series.

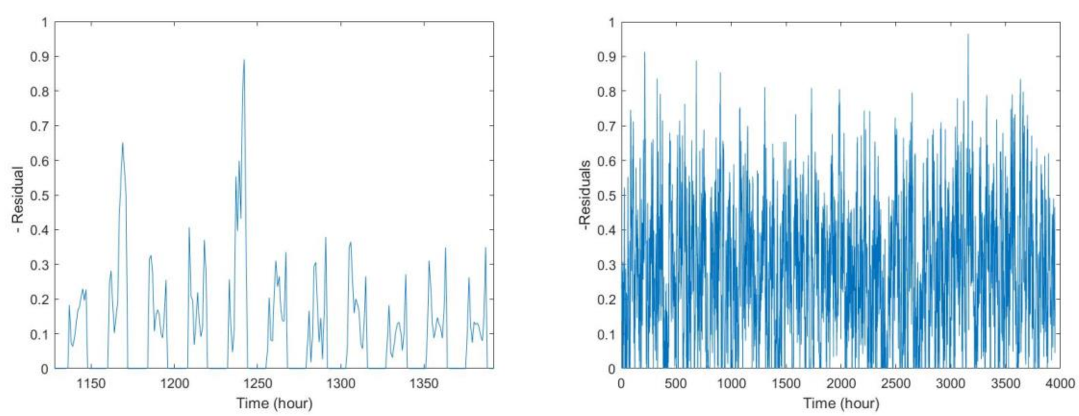

Figure XVI. (a) Detail of raw -Residuals vs time (b) -Residuals vs time for δ=1 (After sunrise, before sundown)

A Dickey and Fuller test was performed to check for stationarity. The test indicated that the hypothesis is TRUE so we can reject the unit root hypothesis with a significance level of pvalue=10^{-5}. Therefore, we can assume that the residual solar energy is stationary.

Next, the sample partial autocorrelation function was obtained to study potential time series models (see Figure XVII). This graphs show a strong autocorrelation with the 1st lag. For simplicity's sake, an AR(1) model was chosen to model the deterministic forecast residuals.

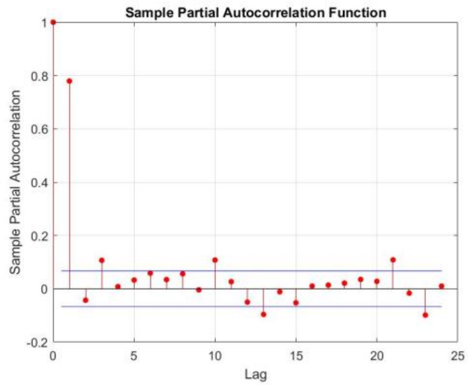

Figure XVII. Sample partial autocorrelation of residual solar energy

Partial autocorrelation function of the time series residuals was obtained as shown in Figure XVIII. This graphs shows that most autocorrelation was removed to a significant level (confidence interval $\pm 1.96\sqrt{M}$ with M=8760 hours).

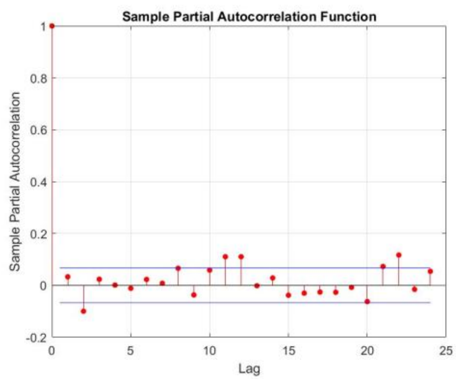

Figure XVIII. Sample partial autocorrelation of residual solar energy

Aggregated Solar Power Energy Simulation

Analogously to wind power energy modelling, a bottom-up simulation can be performed using only the described infrastructure metadata: geographic coordinates provide direct solar irradiation information that can be translated into normalized power using the deterministic regression model according to the capacity (lower or higher than 20MW). Energy output, then, can be calculated using the installed capacity of a solar plant.

Cloud shadowing modelling was simplified by taking a representative AR(1) time series of only one power plant for each capacity category. It is important to outline that this is a simplified approach that does not take into account spatial (eg. a cloud front that shadows nearby PV plants) and temporal correlation.

Solar power park infrastructure by 2026 was extrapolated assuming installed plus to be operational wind power plants will be available by that date (see Figure XIX for example). Given that solar capacity by 2026 will exceed the former projects, additional power plants were added randomly in the main solar project areas following the same distribution as the Renovar projects.

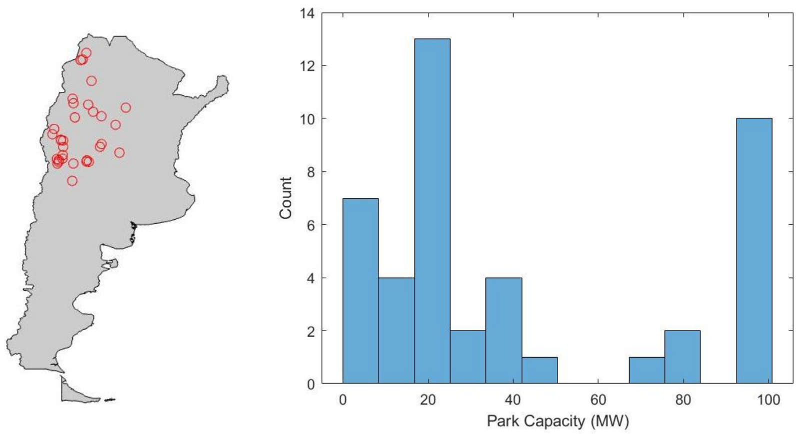

Figure XIX. Distribution of Res.202, Renovar 1, 1.5 and 2 solar park projects

Figure XX shows the result of a simulation using the procedure described in this section.

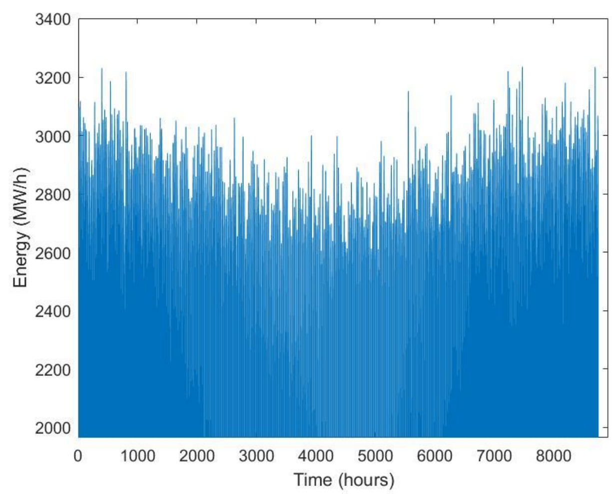

Figure XX. 4,000MW solar power installed capacity energy simulation for 8760h (1year)

Nuclear, Hydro, Biogas and Biomass Energy Generation

Nuclear power infrastructure in Argentina consists of 3 power plants: Embalse (648MW), Atucha I (357MW) and Atucha II (745MW). Capacity by 2025 is very unlikely to change radically given that new power plant projects in Argentina were suspended because of budget constrain, plus the average construction time for a PWR is in average 7.7 years[9]. CAREM 25MW mini nuclear power reactor will be assumed to be operational by 2026 and a power increase for Embalse of 35MW that will be available in 2019. These will be included in the 2026 estimations. Assuming nuclear power generation is flat and an average capacity factor of 85% (eg. Embalse capacity in 2006 was 89% and Atucha capacity in 2016 was 80%) gives a total of 1539MW.
Biomass and biogas power generation in Argentina is rather limited. According to the biomass committee of CADER (Cámara Argentina de Energías Renovables) there are currently between 60 and 80 biomass plants in Argentina of which only 20 are over 1MW (except for Prodeman 10MW peanut biomass plant) [10]. According to CAMMESA January 2019 monthly report, biogas generation at the beginning of 2019 was 24MW. Renovar 1, 1.5 and 2 biomass and biogas awarded projects account for 236MW. Assuming capacity doubles by 2026 to 550MW, energy generation is flat and an average capacity factor of 85%, this gives a total of 465MW.

Current Run-of-the-river hydroelectric power infrastructure in Argentina (Alvarez Condarco, El Tigre, Nihuil IV, Arroyito, Cuesta del Viento, Ullum, Quebrada de Ullum, San Roque, Los Molinos II and Salto Grande) equals 336.5MW plus 945 MW for Salto Grande giving a total of 1281 MW. Mini hydro assigned during Renovar 1, 1.5 and 2 tenders sums 32MW of installed capacity.
Yacyreta hydroelectric plant is an impoundment facility of 3100MW of which 2745MW corresponds to power available for Argentina. Although impoundment facilities have the capacity for storage, Yacyreta is used as a baseload plant keeping dam water levels high which leaves almost no margin for water storage. Therefore, Yacyreta falls into the category of inflexible hydroelectric power generation and should be considered for this calculation.
Projects to be operational by 2026 were identified from which two potential candidates were analyzed for this estimation. The first is a Nahueve run-of-the-river hydroelctric power plant which has 4.6MW of installed capacity. The second was 360MW La Barrancosa (downstream plant of the Santa Cruz hydroelectric project) given an operation constraint realted to an environmental commitment to turbine all stored water within 24 hours.
Taking all the above aspects into consideration, the following was assumed for the estimations. Given that Salto Grande and Yacyretá capacity power adds a significant capacity (3690MW), historical hourly data (years 2011-2018) was gathered to run in the simulations. For the rest of the hydro infrastructure to be operational by 2026, it was assumed that mini hydro capacity triples by 2026 (96MW of installed capacity), energy generation is flat (not a quite realistic assumption) and an capacity factor of 45% (representative of the mode value of a RoR hydro in Argentina), which gives a total of 358MW.

Residual Load Simulations and Results

Based on the methods described in the previous sections, the residual load for Argentina by 2026 was calculated. Scenarios for energy demand (High, Base and Low) were described in the 2026 Scenario Demand Forecast Modelling section. Figure XXI shows 3 simulations ran for each of the scenarios described in Table IV.

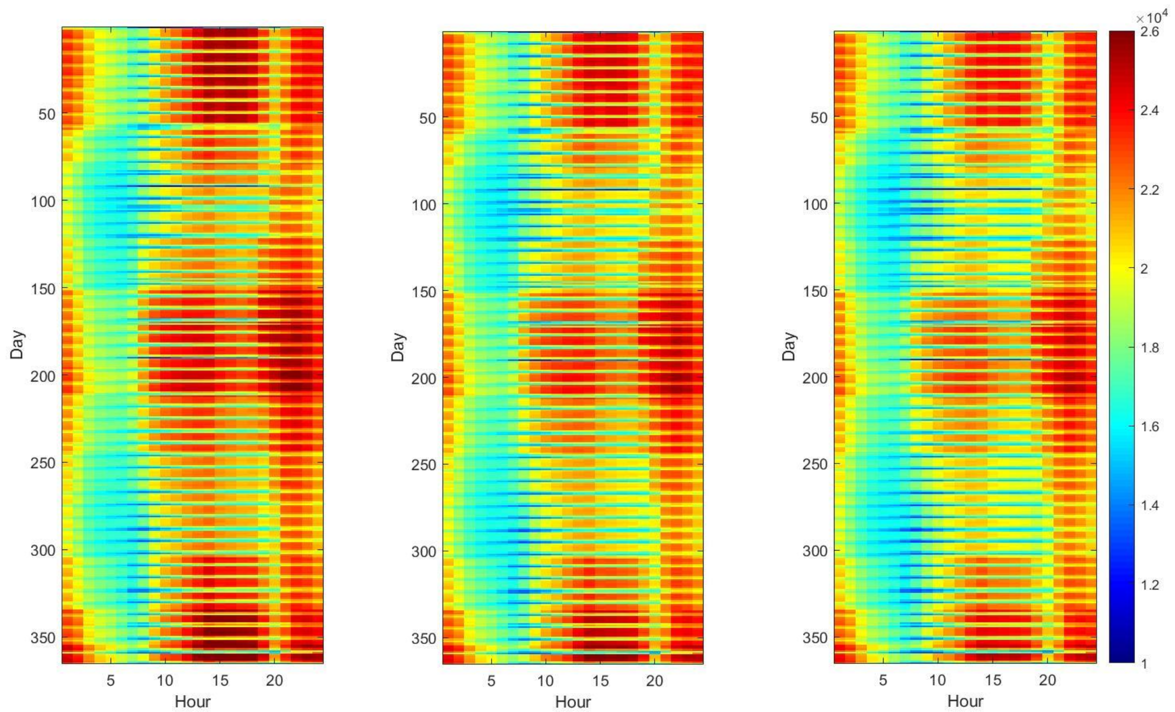

Figure XXI. demand simulations for 2026 scenarios: high (left), base (center), low (right). Graphs show demand differences of up to ~2.000MW. Demand in MW shown in colorscale.

A unique infrastructure scenario with high infrastructure development was assumed with a wind and solar capacity of 10.000MW (60/40 ratio which is representative of the awarded capacity of RenovAr tenders). This gives a total of 16.842MW of must-run+renewable installed capacity as shown in Figure XXII.

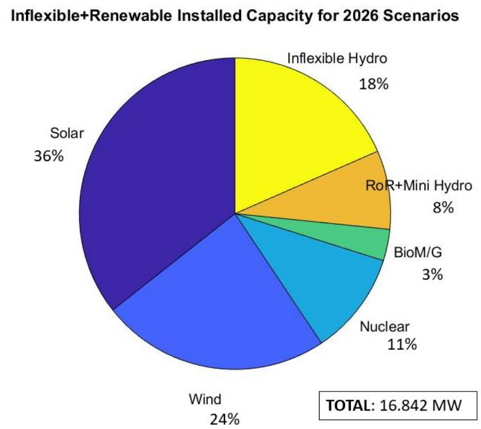

Figure XXII. inflexible and renewable installed capacity for 2026 high infrastructure development scenario

Figure XXIII shows an example of a simulation of inflexible generation in 2026 for the mentioned scenario (taking Year=2012 hydro historic data). It can be observed that daily curves are modulated by solar power. However, peaks do not clearly appear at a specific time (max values predominate before midday in November whereas peaks predominate after midday in January) because of hydro and wind effects.

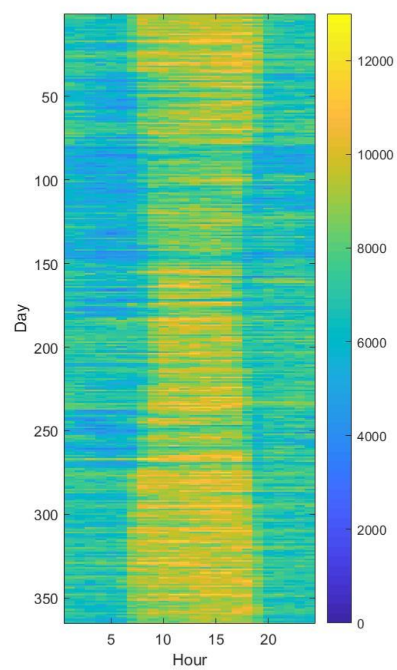

Figure XXIII. Inflexible energy simulation for 2026 scenario. Power in MW shown in colorscale.

Residual load curves were calculated following equation (1) for the 3 mentioned scenarios. These estimations were run 1000 times for each scenario in order to build confidence intervals. Figure XXIII-XXV show the results.

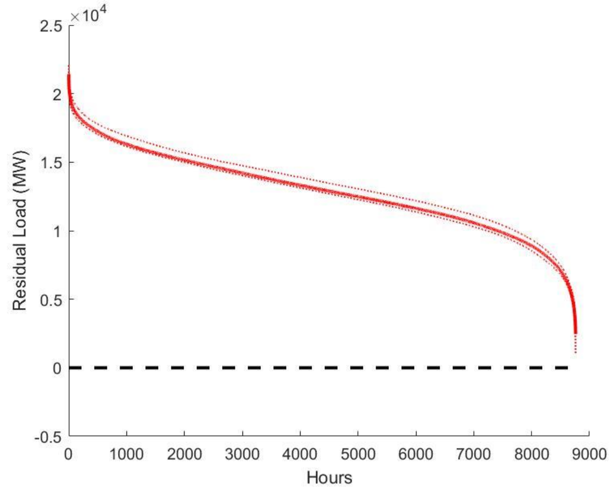

Figure XXIII: residual load curve and 99% confidence intervals for 2026 high demand and high infrastructure development scenario

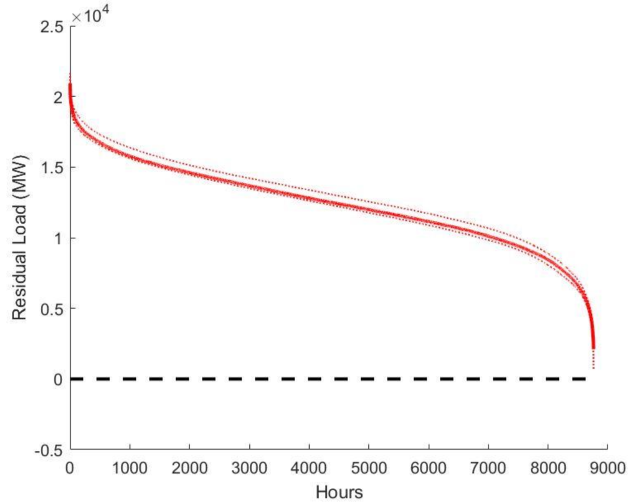

Figure XXIII: residual load curve and 99% confidence intervals for 2026 base demand and high infrastructure development scenario

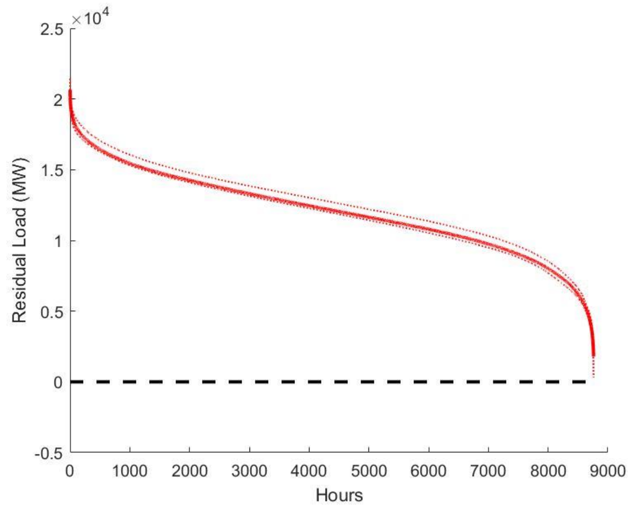

Figure XXIV: residual load curve and 99% confidence intervals for 2026 low demand and high infrastructure development scenario

At a first glance, it can be seen that the residual load curve does not cross below the x axis (i.e. residual load is not negative for any scenario). However, the residual load tail built for the base/low demand scenario approaches values close to 0. The minimum residual load calculated for all experiments was only 79MW –although it is highly unlike to happen-. Figure XXVI shows the result for one simulation (Low Demand Scenario) that helps to visualize the time where the residual load reaches its lowest value. It is worth to outline that this happened (1) during the first days of January (either during midnight or in the morning), (2) in the 18h-20h time interval between April 15th and June 15th, and between September 15^{th} till December, and (3) at midnight during winter holidays.

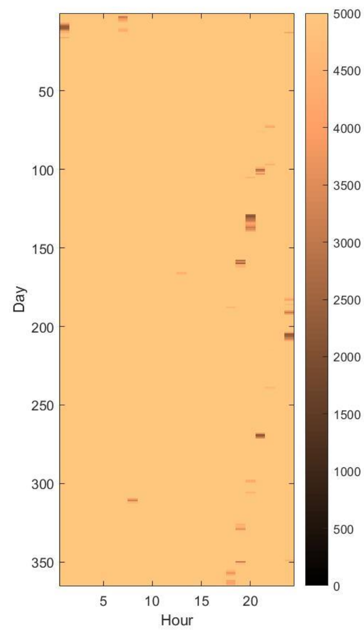

Figure XXVI. residual load simulation for a 2026 low demand scenario. Residual Load in MW shown in colorscale. Max colorscale value was set to 5000MW.

Conclusions

Based on the results described in this paper, residual load is unlikely to become negative by 2026 which means there will not be energy surplus from inflexible generation to be stored, at least, at a moderate scale.
Wind modelling analysis showed correlation between wind parks decays exponentially as distance increases -in accordance with [7]- which can be modelled as $\tau = e^{b*d}$ with b=-4.8*10^{-3}. This aspect should be taken into account when conducting energy planning since it seems Argentina has a trend to cluster wind parks (eg. Bahía Blanca, Arauco, etc.) which could lead to a decrease in residual load.
Solar generation modelling could be also improved given that the current model does not account for spatial and temporal correlation of cloud shadowing. However, this could be included when solar infrastructure deployment makes more information available. The author estimates shadowing correlation is likely to increase extreme values in aggregated solar energy.
Further studies should also include further sensitivity analysis regarding inflexible generation. Wind/solar mix was assumed to be fixed in this analysis. It is not clear what would happen by changing this ratio given that minimum residual load do not occur when solar generation reaches its maximum as it was shown for these simulations.

References

[1]International Energy Agency. *Policies and Measures*. 2018. Web. 29/10/2018. https://www.iea.org/policiesandmeasures/pams/germany/name-34991-en.php

[2] Schill, Wolf-Peter (2013): *Residual load, renewable surplus generation and storage requirements in Germany*, DIW Discussion Papers, No. 1316, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin.

[3] Energy Pedia. *Residual Load. 2019. Web. 18/2/2019.*
https://energypedia.info/wiki/Residual_Load

[4] Berk, Kevin. *Modeling and Forecasting Electricity Demand A Risk Management Perspective*. Springer Fachmedien Wiesbaden GmbH, 2015.

[5] Soliman, Soliman Abdel-Hady, and Ahmad Al Kandari. *Electrical Load Forecasting: Modeling and Model Construction*. Butterworth-Heinemann, 2010.

[6] Gheisa R. T Esteves, Bruno Q. Bastos, Fernando L. Cyrino, Rodrigo F. Calili, Reinaldo C. Souza, *Long Term Electricity Forecast: A Systematic Review, Procedia Computer Science*, Volume 55, 2015, Pages 549-558, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2015.07.041

[7] Malvaldi, A., Weiss, S., Infield, D., Browell, J., Leahy, P., and Foley, A. M. (2017) *A spatial and temporal correlation analysis of aggregate wind power in an ideally interconnected Europe*. *Wind Energ.*, 20: 1315–1329. doi: 10.1002/we.2095.

[8] UCSD SCRIPPS Institute of Oceanogrpahy, Suncycle. 2019. Web. 18/2/2019.
http://mooring.ucsd.edu/software/matlab/doc/toolbox/geo/suncycle.html

[9] Carajilescov, P., Moreira J.M.L. Construction Time of PWRs. 2011 International Nuclear Atlantic Conference - INAC 2011 Belo Horizonte,MG, Brazil, October 24-28, 2011 ASSOCIAÇÃO BRASILEIRA DE ENERGIA NUCLEAR – ABEN.
Available at: https://inis.iaea.org/collection/NCLCollectionStore/_Public/42/105/42105221.pdf

[10] Enel. Crece la Biomasa en Argentina. 2019. Web. 18/2/2019
https://www.enel.com.ar/es/Historias/a201802-biomasa-en-argentina-con-80-plantas.html